○分類学における data-processing と I. P. Ix.*（小野幹雄）　Mikio ONO: Data-processing technique in plant taxonomy

生物分類学の研究者にとって，非常にやっかいな問題である文献処理については，近来非常な発達をとげている機械化されたデータ処理方式（data-processing machin）を導入し過去3世紀にもわたって蓄積された分類学上のぼう大なデータ（taxa の名前，author, publication 等）を正確，びん速に処理しようという考えが1950年頃からぼつぼつ発表されて来た1)2)4)6)7)）。小規模には，各研究者が簡単な hand-sorting panch card に，必要なデータを収録して能率的に整理する程度のものから，Florida の State Plant Board が昆虫について行ったような I. B. M. のパンチカードシステム（P. C. S.）を用いる大規模な例1) まであり事務機械の発達と平行してデータ処理の機械化は各地で実用化の段階に入っている。我国でも金井5) がこの採用を提案しているほか，筆者らの東京都立大学牧野標本館でも，文献ではないが牧野標本の整理と平行して，各標本から収録したデータを24 fields の「ヘリ穴2列孔パンチカードシステム（NBM system）」で処理することを計画し，1962年夏以来実用化に入っている**。

このような data-processing 方式が一般化する場合に一番問題となるのは ‘Coding’ である。Code とは情報やデータを計算機などの機械に伝え，また集計された結果を機械から取り出すためのシンボルで，いわば人間の言葉を機械の言葉に変える約束——辞書のようなものだが個々の研究者や研究機関がそれぞれ全くバラバラな辞書を作ってしまっては不便であり，特に国際的な情報の交換サービスを考えると今のうちに統一的な code を定めておくことが望ましい。

I. B. M. のパンチカードシステム（P. C. S.）のように大規模な機械装置は将来と言えども各研究機関がそれぞれ用意することは不可能なので，どこか中央の機関に設置し各地の研究者の求めに応じてここで集中サービスを行うことが望ましいが，このためにも統一的な code を決めておくことが不可欠の要件である。

この考えのもとに植物界全般にわたって taxa の学名，その author, publication, type 標本の所在とその原産地等の統一的な coding がアメリカの S. W. Gould3) らによって現在すゝめられている。これは米国の National Science Foundation の Office of Science Information Service の grant のもとに，前記の Gould を Director として I. P. I 事務局（International Plant Index Office）が行っている仕事であるが，その最初の業績として “Family Names of the Plant Kingdom” が昨秋刊行された。

§ I. P. Ix. (Family 編)

この Index の第1巻（I. P. Ix.—1 とよぶ）は現在までに発表された植物界全体の family のうち現在一般に妥当とされて使用されているものを ‘Operative families’ と

---

* I. P. Ix.——Data-processing technique in plant taxonomy.

** 牧野標本館で実施しているシステムとその coding については別の機会に報告する。

Tab. 1.

Primary Groups	Orders		Families		Estimated	
	V	Z	V	Z	Genera	Species
V—PLANTAE	252	16	1100	48	17,038	286,528
1—TRACHEOPHYTA						
10 Dicotyledoneae	44		263		8,442	146,581
20 Monocotyledoneae	12		46		2,147	41,714
30 Gymnospermae	5	6	12	18	62	696
40 Pteridophyta	11	10	27	24	384	9,640
5—BRYOPHYTA						
51 Musci	17		83		666	14,252
52 Hepaticae	5		44		259	8,241
53 Anthocerotae	1		2		5	296
6—ALGAE						
61 Rhodophyta	11		57	1	221	2,552
62 Phaeophyta	11		39	1	239	963
63 Pyrrophyta	14		48		96	1,155
64 Chrysophyta	22		90		236	5,453
65 Chlorophyta	11		64		234	4,913
66 Charophyta	1		1	4	6	207
67 Euglenophyta	2		6		14	481
68 Cyanophyta	3		22		118	1,227
7—FUNGI						
71 Phycomycetes	11		39		188	1,060
72 Ascomycetes	16		71		1,393	7,297
73 Basidiomycetes	15		45		483	15,950
74 Fungi-Imperfecti	4		11		1,099	5,714
75—LICHENES						
750 Phycolichenes	1		1		1	1
760 Ascolichenes	12		52		448	16,119
770 Basidiolichenes	1		3		3	8
780 Lichenes-Imperfecti						
8—MYXOPHYTA						
81 Myxomycetae	6		17		57	421
82 Acrasiae	2		5		8	21
83 Plasmodiophoreae	1		1		15	23
85—BACTERIOPHYTA						
850 Schizophyta	11		47		196	1,478
87—VIRA						
870 Vira	2		4		18	65

V=operative names; Z=extinct names.

して，これに4ケタの番号を code し，即に発表されたすべての synonyms を拾ったものがある。また family number の前にやはり4ケタの数字からなる ‘group-indicater’ がついていて，第1のケタは門 (phylum) を現し4ケタ全部で目 (order) まで表示できるようになっている (Tab. 1)。また分類上疑議のある taxon については，その問題とされる rank を数字 ‘9’ で示す。(例えば 6491 Pantostomatineae はこの taxon が 64 =Chrysophyta に属するが Order の rank では異説のあることを示す。)

この group-indicater と family number との間に S〜Z の1文字がはさまれるが，これは ‘kind’ とよび，その taxon (または名称) が現在使用されているもの (前述の operative family) ならば V, synonym ならば S, 化石などの場合は Z を表示する。(この3字以外の T, U, W, X, Y は将来のために保留し，また A〜R は将来動物界についても同様の coding を行う場合，動物に割り当てるために残してある。——寄生，共生などの問題を考えれば動植物全体を cover する coding が望ましい)。したがって ‘kind’ の文字をはさんで前後各4ケタの数字からなる計9ケタの記号 (9 digits の field とよぶ) が family の code となる。この場合，後4 digits の数字だけでも独立に family を表すから表示能力 (capacity) は 9999 で，現在しられるすべての families 1150 の他になお 8850 もの余祐が今後のために残されるわけである。Synonym はその含まれる operative family と同番号で表すから，次のような表示をする。

1662 V 6509 ONAGRACEAE
1662 S 6509 OENOTHERACEAE
1662 S 6509 EPILOBIANAE
1662 S 6509 JUSSIEUACEAE

次に属 (genus) は約 20,000 の operative genera と将来の新属に 80,000 の余裕をのこして 50 digits の field (capacity は 100,000 語) を与える。これもこの field だけで独立に属を表示できるので，使用上の便宜から全体を3群にわけ第1ケタの数字で区別し (Tab. 2), 各群内は alphameric* に code する。

種 (species) については属とつないだ2名法表示をとるので，epithet にだけ 1 field (3 digits) を与え，各属ごとに alphameric に code し，variety についても1ケタの数字をこの後につける。(変種数は10箇しか表示できないが，10以上もの変種があるのは栽培植物で，しかもこの場合はしばしば数十以上に達するから，変種数1〜8の場合はこの field に直接パンチし，9以上のものはすべて ‘9’ で表して，それらには別の field を与える)。

以上全部で 18 digits の記号によりすべて (変種以上) の taxa を code したことに

---

* 属名などを alphabetical に並べて順次に番号を与えて行く coding を alphameric coding という。この code を電子計算機の自動印刷装置 (Compuwriter) にかければ，自動的に alphabet 順の表が印刷される。

Tab. 2.

Genera Number		Primary Groups
First	0 × × × ×	TRACHEOPHYTA
Alpha	≀	
Group	4 × × × ×	
Second	5 × × × ×	BRYOPHYTA or ALGAE
Alpha	≀	
Group	6 × × × ×	
Third	7 × × × ×	FUNGI, LICHENES, MYXOPHYTA BACTERIOPHYTA or VIRA
Alpha	8 × × × ×	
Group	9 × × × ×	

なり，このcodeをguide numberとして，それぞれの学名，author，publication等々を記載したカード(差し当っては各7種)を用意すれば，機械は命令通り必要な種類のカードを高速にえらび出して，指令通りのリスト(学名 ⇄ Author, Author ⇄ Publication, Publication ⇄ 学名等のindex)を自動的に印刷する筈である。

今度I. P. I.が世界に配った‘family names’(I. P. Ix.—1)もこの方法で自動印刷されたものゝコピーである。説明によると基礎cardsの7種類は1) Primary group, 2) Order, 3) Family, 4) Genus, 5) Author, 6) Publication, 7) Speciesで80 columnのI. B. Mカードにパンチしたものである。Taxon以外のauthor, publicationにもそれぞれcodingが必要だが，authorはその姓の頭3字と生年の末尾3ケタ6 digitsでcodeしている。例えば，

Carl Linnaeus (born 1707) は LIN 707
Carl von Linne fil. (born 1741) は LIN 741
Tomitaro Makino (born 1862) は MAK 862

Publicationについては全体を3つに分け，単独の著書D，定期刊行物J，PflanzenfamilienなどのSeriesものはSとして，その後に4ケタの数字と，巻数，ページ，発行年を表示する。Publicationカードはこのcodeから，出版物の名を引きだして印刷するためのものだが，スペースの関係で誌名をフルに載せることはできないので合計34文字以内のabrebiationを定めておく(Biological Abstractなどで用いるものを利用)。この問題は学名カード(species cards)でも同様で，載せる文字は属，種に各17字，計34 digitsのspaceを割り当てている。ほとんどすべての学名はこれで収録できるが，極端に長いものは不可能である。

W. Brownの“Composition of scientific words”によると，もっとも長い学名は

Baikal 湖の Amphipoda で，*Leucophthalmoechinogammarus crassus* とか *Siemienkiewicziechinogammarus siemienkiewiczi* というのがあって後者など実に44文字に達する。こうゆう問題は命名規約の方で制限すべきであろう。

§ I. P. I. の将来の計画

I. P. I. では上記の family name (I. P. Ix.—1) を手始めに，次々に応用的な刊行を計画していて，さしあたり次の6つの index が作られている。

1). Genera, order, class, division 等の monomial names についての alphabetical な一覧表。

2). Vascular plant, mosses, algae, fungi, lichenes, および bacteria についてその種名のリスト。これが全植物界の学名の総索引になるわけで I. P. Ix. の主体となる。

3). Monomial categories の分類体系を示すリスト。

4). Author の index を次の4つの形で作製。即ち，全 author の alphabetical なリスト，慣用的に使われる略名からフルネームを引くリスト，専門とした taxon からの索引，および I. P. Ix. の他のリストやカードに用いている code との対照表。

5). Publication について，単行書の author 別，定期刊行物の誌名別，および series ものの title 別にした，記載 taxa の名との対照 index.

6). Author の完全な bibliography—各 author ごとに，その命名した名と掲載された publication, original の学名が現在どのように扱われているかの対照表。

これらに続いて将来のプランとして31項もの計画があげられているが，主なものは，type 標本の所在とその original 文献のリスト，染色体数，花粉，化石等のデータのリスト，各 taxon の分布——特に広分布種の分布地図，園芸方面への応用のリスト等が予定され，更に，I. P. I. の data cards の duplicate を作って配付することや，将来一般の研究者にカードと設備を開放して応用的研究の場所を提供することまで考えられている。

§問題点

以上，I. P. Ix. の内容をこと細かに説明したのは，これが単なる試みとしてゞはなく，実用化を予期した，積極的な態度でスタートした仕事である以上，早晩世界の standard 化される可能性が強いと思われるからである。その理由は，いかにもアメリカ的な物量と機械化で，きわめて組織的な仕事が進められ，また相当な経済的背景をもって世界への service (若干押し売り的なところも含めて) として行われている点で，恐らく，別の system をもってこれに対抗し得る力は世界のどこにもなさそうに思える。今度の I. P. Ix.—1 を，I. A. P. T. の会員に無償で配布したやり方を見てもこのことは容易に想像できる。とすれば，われわれとしてもこの system についてその内容や欠点をも理解した上で，積極的に利用する態勢をとる方が得策と考えられる。それに早晩，機械的 data-processing の導入が予期される以上，各研究機関でバラバラな code を採用して混乱を生じる前に，多少の批判はあっても (できればその批判を今の中に I. P. I. に提起して

改善を求めた上で）この I. P. Ix. を standard に採用することが望ましい。

この意味で I. P. Ix. について筆者の若干の批判をのべれば：

1. Family に対する Code number が，現在使用されている operative name を alphabetical に並べ，29. Acanthaceae から 9999 Zithiaceae までとびとびの番号を与えている。これは将来の family ために適宜欠番をおいたものであろうが，むしろ現在のものを通し番号で code して，新しい名は順次その後に続けて番号を与えるやり方（いわゆる open-ended system）の方が実際的と思われる。

2. また Family number が全植物界共通の alphametric で与えられているが，これは属同様，3つぐらいの群にわけておく方がよい。牧野標本館では vascular plants を対象として code しているが，I. P. Ix. 方式では 4 digits 要る所をこの方法で 3 digits に節約できる。これは規模の小さい hand-sorting system にとっては切実な問題である。

3. Operative family の名のつけ方が余りに機械的で慣用を無視しすぎるきらいがある。例えば Balanopsidaceae, Capparidaceae がそれぞれ Balanopaceae, Capparaceae となるなど使用上不便である。

4. Synonym の拾いかたが不十分のように思える。特に隠花植物部門でそれが目立つ。例えば我国で普通に使われている Fossombroniaceae（ウロコゴケ科）や Pallaviciniaceae（クモノスゴケ科），Sauteriaceae（アシブトゼニゴケ科）などが synonym としても全く拾われていない。

5. 同様に，当然 operative name となっているはずの Takakiaceae も全く見当らない。

6. Author の code をイニシァル3字と生年とで表わすが，同一年に同一姓の別人が生まれる例は十分あり得る。（分類学者ではないが Cori- エステルで有名な生化学者 C. F. Cori と G. T. Cori とはともに 1896 年プラーハの生まれである。）

7. 植物界の全 taxa をもうらし，その各々について7種（将来は数十種）ものデータカードを用意すると総数は恐らく数百万枚に達する。I. B. M. sorter の能力から言えばこれを処理するのはさして時間を要すまいが，カードの保管だけでも大変である。現在の電子計算システムでは，パンチカードの代りに磁気テープを用いて所要の data を磁性化して store する方式が開発されているので，I. P. I でもこの方式を用いることが望ましい。

8. 将来の問題であるが，各 taxon の分布など地理的データを収録する code として世界全域についての適当な座標表示が必要であろう。牧野標本館では日本列島について，緯度経度とも $^1/_4°$ を単位とする方眼座標（5万分の1地形図と同一区画）を，東西2ケタ，南北3ケタの 5 digits で code し，更にそれを 4 等分した sub-number（1 digit）により 2万5千分の 1 地図と同一区画で表示しているが，世界の地域によって

必要とする精度も異るので，coding の standard について国際的なシムポジアムのような形で討議されることが望ましい。

以上，data-processing machine による分類学 data の処理について，I. P. Ix. の紹介と併わせて論じたが，わが国での具体化について多くの御意見がうかがえれば幸いである。なお I. P. Ix. の批判中，隠花植物部門の引例は都立大学牧野標本館の大西一博氏の御教示によった。

文献

1) Denmark, H. A., H. V. Weems and Carlis Taylor; Taxonomic codification of biological entities. Science **128**: 990-992 (1958).

2. Dysom, G. M.; Indexing scientific progress by computor; New Scientist **49**: 819-820 (1961).

3) International Plant Index; Family names of the plant kingdom. New Haven (1962).

4) Jahn Theodore; Man versus machine, a future problem in protozoan taxonomy: System. Zool. **10** (4): 179-192 (1961).

5) 金井弘夫：植物分類学の方法についての私見. 植物分類地理 **20**: 20-23 (1962).

6) Levine, N. D.; A panched card system for filing parasitological bibliography cards. Journ. Parasitol. **41**: 343-352 (1960).

7) Rogers, David, J., and Taffee, T. Tanimoto; A computer program for classifying plants. Science **132**: 1115-1118 (1960).

8) Sneath, P. H. A; The application of computers to taxonomy. Journ. Gen. Microbial **17**: 201-206 (1957).

Summary

As for the data-processing method in taxonomic works, there are some proposals to apply a punched-card system (P.C.S.) using a computor. One of the most important technical problems in this system, is to make a standard coding. It is desirable that the same coding is used uniformly by all organs which adopt this automatic data-processing method. For this purpose, it has been proposed to establish an international clearing house for coding all biological entities (Denmark et al. 1958). Recently, the International Plant Index wrestles with this massive bussiness under the direction of Dr. Gould, and has published the family index (so-called I. P. Ix.—1) as an initial step of this achievement. This article is an introductry review of this I. P. Ix.—1 with several technical critics for the coding. I hope, more discussion and criticism should be done for both this I. P. Ix. and the system of data-processing itself, and, the adoptation of better coding to contribute for progress of plant taxonomy.